NCGAS: Providing National Cyberinfrastructure to Biologists, esp. Genomicists.

Thomas G. Doak, PI and Manager
National Center for Genome Analysis Support

INDIANA UNIVERSITY

INDIANA UNIVERSITY

Carrie Ganote

Sheri Sanders

Andy Arenson

Bhavya Nalagampalli Papudeshi

Trinity Development
Ben Fulton
Cicada Dennis

Painting of Stylonychia *by Yana Eglit*

NSF-Funded NCGAS partner
Access to Bridges for big big jobs
Lead on Microbiome projects
Partner on GenePattern hosting
Lead in XSEDE involvement

PSC
PITTSBURGH SUPERCOMPUTING CENTER

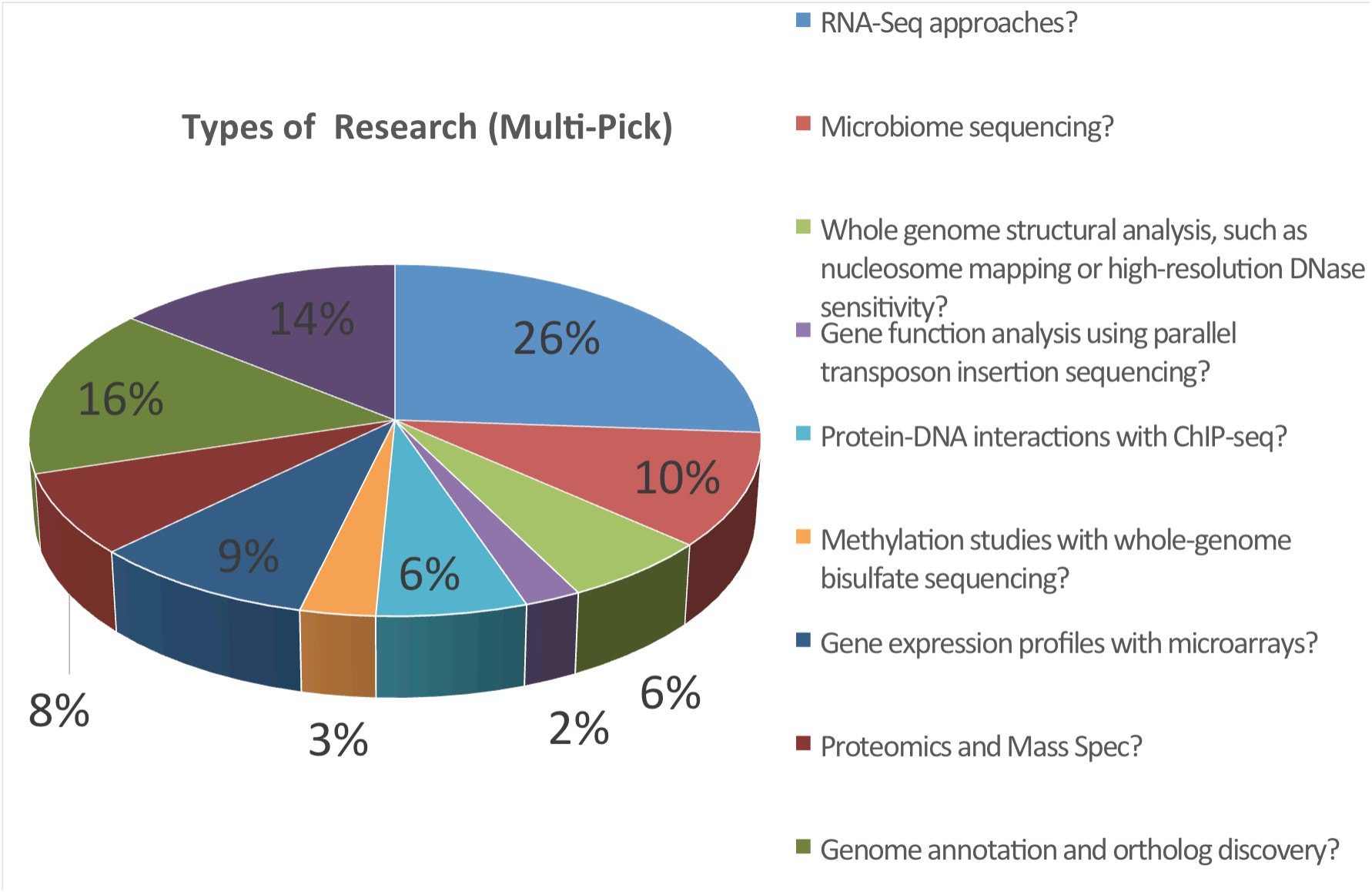
Supporting NCGAS Genomics Research at PSC

Philip D. Blood, Ph.D.
Senior Computational Scientist
Pittsburgh Supercomputing Center

Why Genomics?

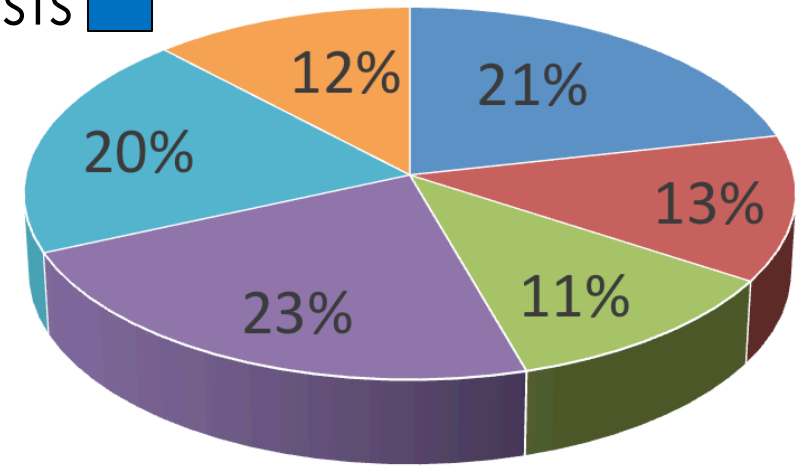
- Genomics, transcriptomics, etc. 'omics have and are revolutionizing everything from ecology to cancer treatment.
- Genomics is computationally demanding
- There are many new researchers, whose specialties lie elsewhere
- IU has a rich history in genomics, and RT in helping them.
- NSF agreed

From our NSF survey:



NCGAS's primary goals:

- Provide bioinformatics expertise through consultation
- Maintain a curated set of genomics applications
- Provide access to HPC resources, esp. large-memory clusters, e.g. Mason, Bridges, Jetstream
- Build Galaxy instances for our software
- Pursue outreach to biologists



Who NCGAS serves and how:

- IU/IUPUI students and researchers
- NSF researchers nation-wide as an NSF-funded service center
- Cancer researchers (ITCR) as a partner on grants
- "For pay" work on others' grants (NSF, USDA, *etc*.)

Environmental Genomics Workshop 2016

MDI Biological Laboratory

MDI Biological Laboratory hosted Environmental Genomics in Salisbury Cove, Maine. Over one week, 9 students and 9 post doc/faculty from 14 universities learned to design RNA-seq experiments, create Daphnia RNA-seq libraries, manage and analyze the data, and present their results.

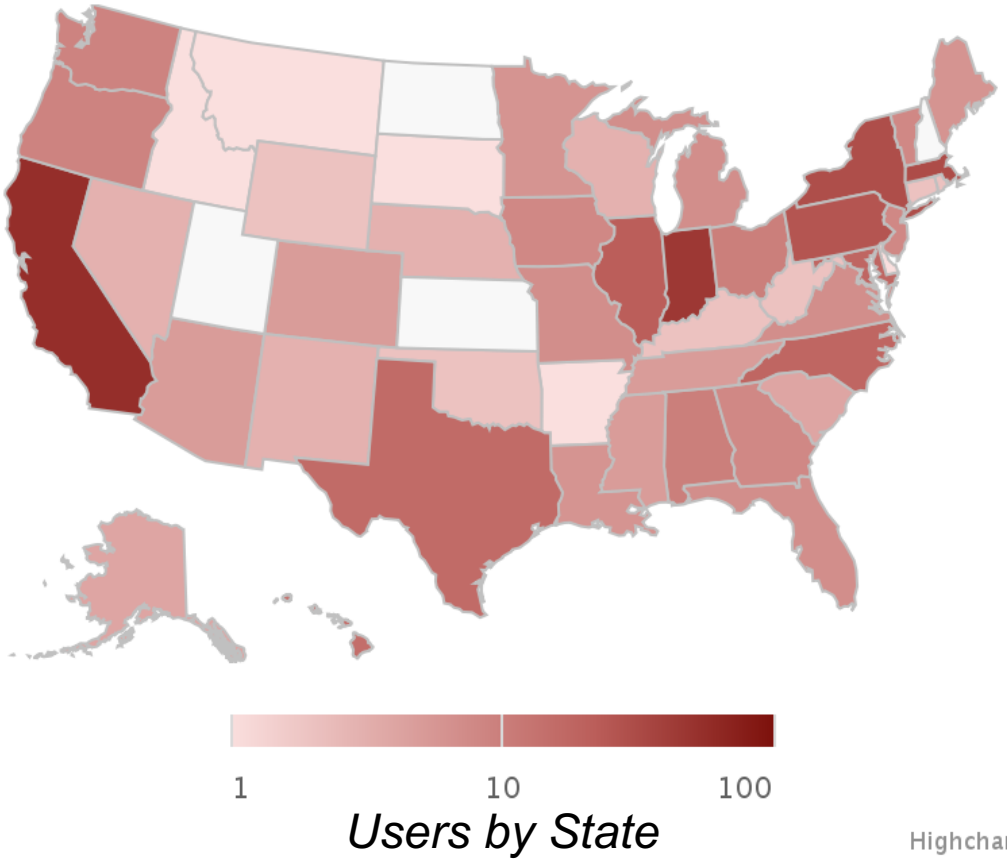
***NCGAS** partnered with the workshop and provided:*

- *Three reserved Karst nodes, totaling*
  - *48 processors*
  - *48 GBs of RAM*
- *On site training and consultation on cluster use, bioinformatics, and statistics. Immediate coordination with IU sysadmin.*
- *Rapid processing of ~250Gb of sequence data, when data arrival was delayed*

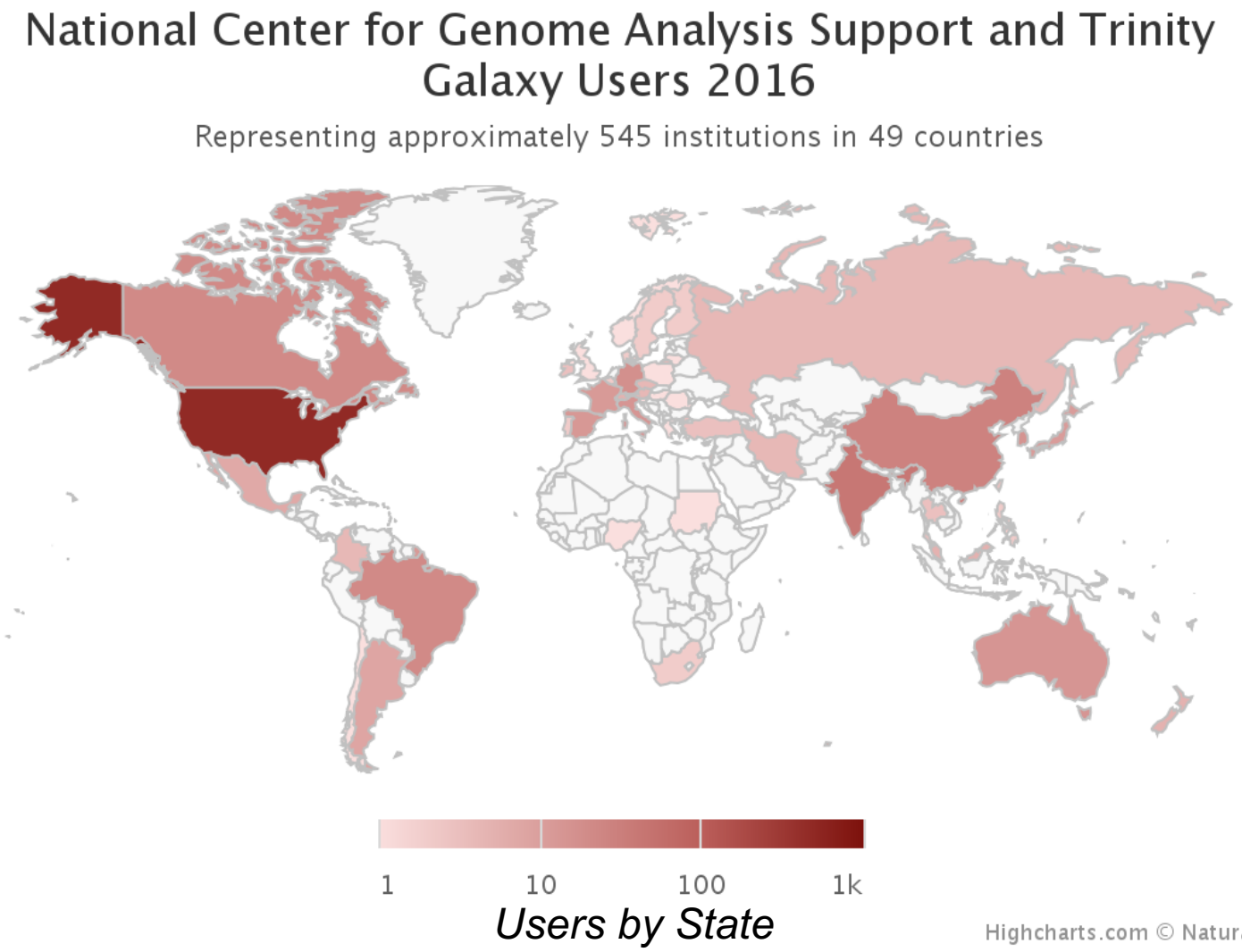
As a result of NCGAS partnership, the workshop was able to include 4.5x more data throughput, enabling more realistic and complex experimental designs.

National Center for Genome Analysis Support and Trinity Galaxy Users 2016

Representing approximately 233 institutions



Users by State


National Center for Genome Analysis Support and Trinity Galaxy Users 2016
Representing approximately 545 institutions in 49 countries
1
10
100
1k
Users by State
Highcharts.com © Natural Earth

Carrie Ganote

Sheri Sanders

Andy Arenson

Bhavya Nalagampalli Papudeshi

Trinity Development

Ben Fulton

Cicada Dennis

Painting of Stylonychia *by Yana Eglit*